

PHYTYL/PRENYLTRANSFERASE NUCLEIC ACIDS, POLYPEPTIDES
AND USES THEREOF

Dean Della Penna
 Eva Collakova
 Sean J. Coughlan
 Timothy G. Helentjaris

CROSS-REFERENCE TO RELATED APPLICATIONS

This application ^{is a CIP} ~~claims benefit~~ of U.S. Patent Application Serial No. 09/307,460 filed May 7, 1999, ^{now abandoned} which is hereby incorporated by reference.

TECHNICAL FIELD

The present invention relates generally to plant molecular biology. More specifically, it relates to nucleic acids and methods for modulating their expression in plants.

BACKGROUND OF THE INVENTION

The chloroplasts of higher plants contain and elaborate many unique, interconnected biochemical pathways that produce an array of compounds that not only perform vital plastid functions but are also important from agricultural and nutritional perspectives. One class of lipid soluble, chloroplastically synthesized compounds are the prenyllipids, plastoquinone and tocopherols. Plastoquinone is a fundamentally important component of the chloroplast photosynthetic electron transport chain and accounts for up to 50% of the total prenyllipid pool in green tissues. Tocopherols collectively account for up to 40% of the total prenyllipids pool in green plastids and have a well documented role in mammals as an antioxidant [Liebler, 1993] and a similar, though less well understood antioxidant role in plants [Hess, 1993]. The essential nutritional value of tocopherols has been known for over 70 years [Mason, 1980]. Despite the well studied, wide-spread importance of these chloroplastic compounds to human nutrition, agriculture and biochemical processes within plant cells, much remains to be learned at the molecular level about their biosynthesis.

Plastoquinone and tocopherols are the most abundant prenyllipids in the plastid and are synthesized by the common pathway reviewed in Hess, 1993 and Soll, 1987. The head group for both compounds, homogentisic acid, is produced from p-hydroxyphenylpyruvic acid by the enzyme p-hydroxyphenylpyruvic acid dioxygenase in a reaction that catalyzes both an oxidation and decarboxylation. Homogentisic acid is subject to phytylation/prenylation (phytyl and solanyl, C20 and C45, respectively) coupled to a simultaneous decarboxylation to form the first true tocopherol and plastoquinone intermediates, 2-demethyl-phytylplastoquinol and 2-demethylplastoquinol-9, respectively. A single ring methylation occurs on 2-demethylplastoquinol to yield plastoquinol-9 that is then oxidized to plastoquinone-9. The preferred route in spinach for α -tocopherol formation is thought to be 1) ring methylation of 2-demethylphytylplastoquinol, to yield phytylplastoquinol, 2) cyclization to yield gamma-tocopherol and, finally, 3) a second ring methylation to yield α -tocopherol. The first ring methylation in both tocopherol and plastoquinone synthesis is thought to be carried out by a single enzyme that is specific for the site of methylation on the ring but has broad substrate specificity and accommodates both classes of compounds. The final methylation enzyme (gamma tocopherol methyl transferase) is the only enzyme of the pathway that has been purified from plants to date (dHarlingue and Camara, 1985). All other enzymatic activities of tocopherol/plastoquinone synthesis have been localized to the inner chloroplast envelope by fractionation studies except p-hydroxyphenylpyruvic acid dioxygenase and the tocopherol cyclase enzyme. Difficulties with cell fractionation methods, low activities for some of the enzymes, substrate stability and availability and assay problems make studying the pathway biochemically extremely challenging.

The fact that tocopherol and plastoquinone levels, ratios and total amounts vary by orders of magnitude in different plant tissues and developmental stages indicates the pathway is both highly regulated and highly flexible and has potential for quantitative and qualitative manipulation. However, while biochemical analysis has been useful in deciphering the biosynthetic pathway such studies have provided almost no insight into how bulk carbon flow through the pathway is regulated or how differing amounts of tocopherols or plastoquinone are synthesized.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide nucleic acids and polypeptides relating to the biosynthesis of tocopherol and plastiquinone.

It is another object of the present invention to provide nucleic acids and polypeptides that can be used to identify proteins involved in tocopherol and plastiquinone biosynthesis.

It is another object of the present invention to provide antigenic fragments of the polypeptides of the present invention.

It is another object of the present invention to provide transgenic plants comprising the nucleic acids of the present invention.

It is another object of the present invention to provide methods for modulating, in a transgenic plant, the expression of the nucleic acids of the present invention.

It is another object of the present invention to provide a method for modulating the level of tocopherols and plastiquinone in a plant.

Other aspects of the present invention include expression cassettes comprising the nucleic acid operably linked to a promoter, host cells transfected with the expression cassette, and transgenic plants and seeds comprising the expression cassette.

In a further aspect, the present invention relates to a method of modulating expression of the nucleic acids in a plant, comprising the steps of

- (a) transforming a plant cell with an expression cassette comprising a nucleic acid of the present invention operably linked to a promoter in sense or antisense orientation;
- (b) growing the plant cell under plant growing conditions to produce a regenerated plant capable of expressing the nucleic acid for a time sufficient to modulate expression of the nucleic acids in the plant compared to a corresponding non-transformed plant.

Expression of the nucleic acids encoding the proteins of the present invention can be increased or decreased relative to a non-transformed control plant.

DETAILED DESCRIPTION OF THE INVENTION

Tocopherols are synthesized in the inner plastid membrane. The first committed step in the pathway is the condensation of the homogentisate head group with the phytyl tail catalyzed by an integral membrane protein: homogentisate: phytyl transferase. The present polypeptides catalyze the condensation of homogentisic acid with phytyldiphosphate or geranylgeranyl pyrophosphate to produce the first intermediates in tocopherol or tocotrienol synthesis, respectively.

The phytylation/prenylation of homogentisic acid is a likely key regulatory step for "tail" synthesis and in determining the relative amounts of tocopherols, tocotrienols and plastoquinone produced as it is the branchpoint for the tocopherol and plastoquinone arms of the pathway.

One purpose of this invention is to modulate a prenyllipid biosynthetic pathway, such as the plastoquinone and tocopherol pathways. The modulation of the pathway may be an up regulation or down regulation of the amount or activity of a prenyllipid (ie. plastoquinone or tocopherol), or of an intermediate in a pathway (ie. 2-demethyl-phytylplastoquinol or 2-demethylplastoquinol-9).

DEFINITIONS

The term "isolated" refers to material, such as a nucleic acid or a protein, which is: (1) substantially or essentially free from components which normally accompany or interact with the material as found in its naturally occurring environment or (2) if the material is in its natural environment, the material has been altered by deliberate human intervention to a composition and/or placed at a locus in the cell other than the locus native to the material.

The terms polypeptide, "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The essential nature of such analogues of naturally occurring amino acids is that, when incorporated into a protein, that protein is specifically reactive to antibodies elicited to the same protein but consisting entirely of naturally occurring amino acids. The terms "polypeptide", "peptide" and "protein" are also inclusive of modifications including, but not limited

to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation. Further, this invention contemplates the use of both the methionine-containing and the methionine-less amino terminal variants of the protein of the invention.

As used herein, "plant" includes but is not limited to plant cells, plant tissue and plant seeds.

As used herein, "promoter" includes reference to a region of DNA upstream from the start of transcription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription.

By "fragment" is intended a portion of the nucleotide sequence or a portion of the amino acid sequence and hence protein encoded thereby. Preferably fragments of a nucleotide sequence may encode protein fragments that retain the biological activity of the native nucleic acid. However, fragments of a nucleotide sequence which are useful as hybridization probes generally do not encode fragment proteins retaining biological activity. Fragments of a nucleotide sequence are generally greater than 10 nucleotides, preferably at least 20 nucleotides and up to the entire nucleotide sequence encoding the proteins of the invention. Generally probes are less than 1000 nucleotides and preferably less than 500 nucleotides. Fragments of the invention include antisense sequences used to decrease expression of the inventive nucleic acids. Such antisense fragments may vary in length ranging from at least about 20 nucleotides, about 50 nucleotides, about 100 nucleotides, up to and including the entire coding sequence.

By "variants" is intended substantially similar sequences. Generally, nucleic acid sequence variants of the invention will have at least 50%, 60%, 70%, or preferably 80%, more preferably at least 90% and most preferably at least 95% sequence identity to the native nucleotide sequence.

Generally, polypeptide sequence variants of the invention will have at least about 55%, 60%, 70%, 80%, or preferably at least about 90% and more preferably at least about 95% sequence identity to the native protein.

As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in

be 100% if the 40 nucleotides or amino acids of the shorter sequence match 40 nucleotides or amino acids of the larger sequence. This is because GAP may calculate the percent identity based on the total length of the shorter sequence. However, where this specification, including the claims, specifies the sequence identity being computed by GAP, the GAP percentage identity should be recalculated as a percentage of the longer sequence and any nucleotides or amino acids in the larger sequence that extend beyond the shorter sequence would not count as a match. In the example provided above this would give a percent identity of 40%.

Other methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, *Adv. Appl. Math.* 2:482 (1981); by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48:443 (1970); by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci.* 85:2444 (1988); by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California; GAP, BESTFIT, FASTA, BLAST and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wisconsin, USA; the CLUSTAL program is well described by Higgins and Sharp, *Gene* 73:237-244 (1988); Higgins and Sharp, *CABIOS* 5:151-153 (1989); Corpet *et al.*, *Nucleic Acids Research* 16:10881-90 (1988); Huang *et al.*, *Computer Applications in the Biosciences* 8:155-65 (1992), and Pearson *et al.*, *Methods in Molecular Biology* 24:307-331 (1994).

The BLAST family of programs which can be used for database similarity searches includes: BLASTN for nucleotide query sequences against nucleotide database sequences; BLASTX for nucleotide query sequences against protein database sequences; BLASTP for protein query sequences against protein database sequences; TBLASTN for protein query sequences against nucleotide database sequences; and TBLASTX for nucleotide query sequences against nucleotide database sequences. See *Current Protocols in Molecular Biology*, Chapter 19, Ausubel *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). The BLAST algorithm performs a statistical

analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5877 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance.

The term "functional equivalent" means that the sequence of the variant polynucleotide defines a chain that produces a protein having substantially the same biological effect as the protein encoded by the non-variant polynucleotide.

The term "Complement" or "Complementary" when used with respect to a polynucleotide sequence refers to the corresponding base pairs in the same sequence.

The term "Hybridization Probe" refers to a process whereby a polynucleotide is used to find a complementary polynucleotide through the annealing of the two polynucleotides to form a double helix.

The term "Coding Sequence" when used with respect to a complete gene sequence refers to the sequence spanning the start and stop codon, and when used with respect to a partial gene sequence refers to a portion of the coding region spanning the start and stop codon.

NUCLEIC ACIDS

The isolated nucleic acids of the present invention can be made using (a) standard recombinant methods, (b) synthetic techniques, or combinations thereof. In some embodiments, the polynucleotides of the present invention will be cloned, amplified, or otherwise constructed from a monocot or dicot. In preferred embodiments the monocot is corn, sorghum, barley, wheat, millet, or rice. Preferred dicots include soybeans, sunflower, canola, alfalfa, cotton, potato, cassava, *Arabidopsis thaliana*, tomato, *Brassica* vegetables, peppers, potatoes, apples, spinach, or lettuce.

Functional fragments included in the invention can be obtained using primers that selectively hybridize under stringent conditions. Primers are generally at least 12 bases in length and can be as high as 200 bases, but will generally be from 15 to 75, preferably from 15 to 50. Functional fragments can be identified using a variety of techniques such as restriction analysis, Southern analysis, primer extension analysis, and DNA sequence analysis.

The present invention includes a plurality of polynucleotides that encode for the identical amino acid sequence. The degeneracy of the genetic code allows for such "silent variations" which can be used, for example, to selectively hybridize and detect allelic variants of polynucleotides of the present invention. Additionally, the present invention includes isolated nucleic acids comprising allelic variants. The term "allele" as used herein refers to a related nucleic acid of the same gene.

Variants of nucleic acids included in the invention can be obtained, for example, by oligonucleotide-directed mutagenesis, linker-scanning mutagenesis, mutagenesis using the polymerase chain reaction, and the like. See, for example, Ausubel, pages 8.0.3 - 8.5.9. Also, see generally, McPherson (ed.), *DIRECTED MUTAGENESIS: A Practical approach*, (IRL Press, 1991). Thus, the present invention also encompasses DNA molecules comprising nucleotide sequences that have substantial sequence identity with the inventive sequences.

Variants included in the invention may contain individual substitutions, deletions or additions to the nucleic acid or polypeptide sequences. Such changes will alter, add or delete a single amino acid or a small percentage of amino acids in the encoded sequence. Variants are referred to as "conservatively modified variants" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. When the nucleic acid is prepared or altered synthetically, advantage can be taken of known codon preferences of the intended host.

The present invention also includes "shufflents" produced by sequence shuffling of the inventive polynucleotides to obtain a desired characteristic. Sequence shuffling is described in PCT publication No. 96/19256. See also, Zhang, J. H., *et al.*, *Proc. Natl. Acad. Sci. USA* 94:4504-4509 (1997).

The present invention also includes the use of 5' and/or 3' UTR regions for modulation of translation of heterologous coding sequences. Positive sequence motifs include translational initiation consensus sequences (Kozak, *Nucleic Acids Res.* 15:8125 (1987)) and the 7-methylguanosine cap structure (Drummond *et al.*, *Nucleic Acids Res.* 13:7375 (1985)). Negative elements include stable intramolecular 5' UTR stem-loop structures (Muesing *et al.*, *Cell* 48:691 (1987)) and AUG sequences or short open reading frames preceded by an appropriate AUG in the 5' UTR (Kozak, *supra*, Rao *et al.*, *Mol. and Cell. Biol.* 8:284 (1988)).

Further, the polypeptide-encoding segments of the polynucleotides of the present invention can be modified to alter codon usage. Altered codon usage can be employed to alter translational efficiency and/or to optimize the coding sequence for expression in a desired host or to optimize the codon usage in a heterologous sequence for expression in maize. Codon usage in the coding regions of the polynucleotides of the present invention can be analyzed statistically using commercially available software packages such as "Codon Preference" available from the University of Wisconsin Genetics Computer Group (see Devereaux *et al.*, *Nucleic Acids Res.* 12:387-395 (1984)) or MacVector 4.1 (Eastman Kodak Co., New Haven, Conn.).

For example, the inventive nucleic acids can be optimized for enhanced expression in organisms of interest. See, for example, EPA0359472; WO91/16432; Perlak *et al.*, *Proc. Natl. Acad. Sci. USA* 88:3324-3328 (1991); and Murray *et al.*, *Nucleic Acids Res.* 17:477-498 (1989). In this manner, the genes can be synthesized utilizing species-preferred codons. See, for example, Murray *et al.*, *Nucleic Acids Res.* 17:477-498 (1989), the disclosure of which is incorporated herein by reference.

The present invention provides subsequences comprising isolated nucleic acids containing at least 16 contiguous bases of the inventive sequences. For example the isolated nucleic acid includes those comprising at least 20, 25, 30, 40, 50, 60, 75 or 100 or more contiguous nucleotides of the inventive sequences. Subsequences of the isolated nucleic acid can be used to modulate or detect gene expression by introducing into the subsequences compounds which bind, intercalate, cleave and/or crosslink to nucleic acids.

The nucleic acids of the invention may conveniently comprise a multi-cloning site comprising one or more endonuclease restriction sites inserted into the nucleic acid to aid in isolation of the polynucleotide. Also, translatable sequences may be inserted to aid in the isolation of the translated polynucleotide of the present invention. For example, a hexa-histidine marker sequence provides a convenient means to purify the proteins of the present invention.

A polynucleotide of the present invention can be attached to a vector, adapter, promoter, transit peptide or linker for cloning and/or expression of a polynucleotide of the present invention. Additional sequences may be added to such cloning and/or expression sequences to optimize their function in cloning

and/or expression, to aid in isolation of the polynucleotide, or to improve the introduction of the polynucleotide into a cell. Use of cloning vectors, expression vectors, adapters, and linkers is well known and extensively described in the art. For a description of such nucleic acids see, for example, Stratagene Cloning Systems, Catalogs 1995, 1996, 1997 (La Jolla, CA); and, Amersham Life Sciences, Inc, Catalog '97 (Arlington Heights, IL).

The isolated nucleic acid compositions of this invention, such as RNA, cDNA, genomic DNA, or a hybrid thereof, can be obtained from plant biological sources using any number of cloning methodologies known to those of skill in the art. In some embodiments, oligonucleotide probes that selectively hybridize, under stringent conditions, to the polynucleotides of the present invention are used to identify the desired sequence in a cDNA or genomic DNA library.

Exemplary total RNA and mRNA isolation protocols are described in *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); and, *Current Protocols in Molecular Biology*, Ausubel, et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995). Total RNA and mRNA isolation kits are commercially available from vendors such as Stratagene (La Jolla, CA), Clontech (Palo Alto, CA), Pharmacia (Piscataway, NJ), and 5'-3' (Paoli, PA). See also, U.S. Patent Nos. 5,614,391; and, 5,459,253.

Typical cDNA synthesis protocols are well known to the skilled artisan and are described in such standard references as: *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); and, *Current Protocols in Molecular Biology*, Ausubel, et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995). cDNA synthesis kits are available from a variety of commercial vendors such as Stratagene or Pharmacia.

An exemplary method of constructing a greater than 95% pure full-length cDNA library is described by Carninci et al., *Genomics* 37:327-336 (1996). Other methods for producing full-length libraries are known in the art. See, e.g., Edery et al., *Mol. Cell Biol.* 15(6):3363-3371 (1995); and, PCT Application WO 96/34981.

It is often convenient to normalize a cDNA library to create a library in which each clone is more equally represented. A number of approaches to normalize cDNA libraries are known in the art. Construction of normalized libraries is described in Ko, *Nucl. Acids. Res.* 18(19):5705-5711 (1990); Patanjali et al., *Proc.*

Natl. Acad. USA 88:1943-1947 (1991); U.S. Patents 5,482,685 and 5,637,685; and Soares *et al.*, *Proc. Natl. Acad. Sci. USA* 91:9228-9232 (1994).

Subtracted cDNA libraries are another means to increase the proportion of less abundant cDNA species. See, Foote *et al.* in, *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); Kho and Zarbl, *Technique*, 3(2):58-63 (1991); Sive and St. John, *Nucl. Acids Res.* 16(22):10937 (1988); *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); and, Swaroop *et al.*, *Nucl. Acids Res.* 19(8):1954 (1991). cDNA subtraction kits are commercially available. See, e.g., PCR-Select (Clontech).

To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation. Examples of appropriate molecular biological techniques and instructions are found in Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Vols. 1-3 (1989), *Methods in Enzymology*, Vol. 152: *Guide to Molecular Cloning Techniques*, Berger and Kimmel, Eds., San Diego: Academic Press, Inc. (1987), *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997). Kits for construction of genomic libraries are also commercially available.

The cDNA or genomic library can be screened using a probe based upon the sequence of a nucleic acid of the present invention such as those disclosed herein. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Those of skill in the art will appreciate that various degrees of stringency of hybridization can be employed in the assay; and either the hybridization or the wash medium can be stringent. The degree of stringency can be controlled by temperature, ionic strength, pH and the presence of a partially denaturing solvent such as formamide.

Typically, stringent hybridization conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long

probes (e.g., greater than 50 nucleotides). Typically the hybridization will be conducted for about 4 to about 12 hours.

Preferably the hybridization is conducted under low stringency conditions which include hybridization with a buffer solution of 30 % formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulfate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50°C. More preferably the hybridization is conducted under moderate stringency conditions which include hybridization in 40 % formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55°C. Most preferably the hybridization is conducted under high stringency conditions which include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60°C.

An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993); and *Current Protocols in Molecular Biology*, Chapter 2, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995). Often, cDNA libraries will be normalized to increase the representation of relatively rare cDNAs.

The nucleic acids of the invention can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of polynucleotides of the present invention and related genes directly from genomic DNA or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes.

Examples of techniques useful for *in vitro* amplification methods are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, U.S. Patent No. 4,683,202 (1987); and, *PCR Protocols A Guide to Methods and Applications*, Innis *et al.*, Eds., Academic Press Inc., San Diego, CA (1990). Commercially available kits for genomic PCR amplification are known in the art. See, e.g., Advantage-GC Genomic PCR Kit (Clontech). The T4 gene 32 protein (Boehringer Mannheim) can be used to improve yield of long PCR products.

PCR-based screening methods have also been described. Wilfinger *et al.* describe a PCR-based method in which the longest cDNA is identified in the first step so that incomplete clones can be eliminated from study. *BioTechniques*, 22(3): 481-486 (1997).

In one aspect of the invention, nucleic acids can be amplified from a plant nucleic acid library. The nucleic acid library may be a cDNA library, a genomic library, or a library generally constructed from nuclear transcripts at any stage of intron processing. Libraries can be made from a variety of plant tissues.

Alternatively, the sequences of the invention can be used to isolate corresponding sequences in other organisms, particularly other plants, more particularly, other monocots. In this manner, methods such as PCR, hybridization, and the like can be used to identify such sequences having substantial sequence identity to the sequences of the invention. See, for example, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York). and Innis *et al.* (1990), *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York). Coding sequences isolated based on their sequence identity to the entire inventive coding sequences set forth herein or to fragments thereof are encompassed by the present invention.

The isolated nucleic acids of the present invention can also be prepared by direct chemical synthesis by methods such as the phosphotriester method of Narang *et al.*, *Meth. Enzymol.* 68:90-99 (1979); the phosphodiester method of Brown *et al.*, *Meth. Enzymol.* 68:109-151 (1979); the diethylphosphoramidite method of Beaucage *et al.*, *Tetra. Lett.* 22:1859-1862 (1981); the solid phase phosphoramidite triester method described by Beaucage and Caruthers, *Tetra. Letts.* 22(20):1859-1862 (1981), e.g., using an automated synthesizer, e.g., as described in Needham-VanDevanter *et al.*, *Nucleic Acids Res.*, 12:6159-6168 (1984); and, the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis generally produces a single stranded oligonucleotide. This may be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill will recognize that while chemical synthesis of DNA is limited to sequences of about 100 bases, longer sequences may be obtained by the ligation of shorter sequences.

EXPRESSION CASSETTES

In another embodiment expression cassettes comprising isolated nucleic acids of the present invention are provided. An expression cassette will typically comprise a polynucleotide of the present invention operably linked to transcriptional initiation regulatory sequences which will direct the transcription of the polynucleotide in the intended host cell, such as tissues of a transformed plant.

The construction of expression cassettes that can be employed in conjunction with the present invention is well known to those of skill in the art in light of the present disclosure. See, e.g., Sambrook, *et al.*; Molecular Cloning: A Laboratory Manual; Cold Spring Harbor, New York; (1989); Gelvin, *et al.*; Plant Molecular Biology Manual; (1990); Plant Biotechnology: Commercial Prospects and Problems, eds. Prakash, *et al.*; Oxford & IBH Publishing Co.; New Delhi, India; (1993); and Heslot, *et al.*; Molecular Biology and Genetic Engineering of Yeasts; CRC Press, Inc., USA; (1992); each incorporated herein in its entirety by reference.

For example, plant expression vectors may include (1) a cloned plant nucleic acid under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (e.g., one conferring inducible, constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific/selective expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Constitutive, tissue-preferred or inducible promoters can be employed. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the ubiquitin 1 promoter, the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Patent No. 5,683,439), the *Nos* promoter, the pEmu promoter, the rubisco promoter, the GRP1-8 promoter and other transcription initiation regions from various plant genes known to those of skill.

Examples of inducible promoters are the Adh1 promoter which is inducible by hypoxia or cold stress, the Hsp70 promoter which is inducible by heat stress,

and the PPK promoter which is inducible by light. Also useful are promoters which are chemically inducible.

Examples of promoters under developmental control include promoters that initiate transcription preferentially in certain tissues, such as leaves, roots, fruit, seeds, or flowers. An exemplary promoter is the anther specific promoter 5126 (U.S. Patent Nos. 5,689,049 and 5,689,051). Examples of seed-preferred promoters include, but are not limited to, 27 kD gamma zein promoter and waxy promoter, Boronat,A., Martinez,M.C., Reina,M., Puigdomenech,P. and Palau,J.; Isolation and sequencing of a 28 kD glutelin-2 gene from maize: Common elements in the 5' flanking regions among zein and glutelin genes; Plant Sci. 47, 95-102 (1986) and Reina,M., Ponte,I., Guillen,P., Boronat,A. and Palau,J., Sequence analysis of a genomic clone encoding a Zc2 protein from Zea mays W64 A, Nucleic Acids Res. 18 (21), 6426 (1990). See the following site relating to the waxy promoter: Kloesgen,R.B., Gierl,A., Schwarz-Sommer,ZS. and Saedler,H., Molecular analysis of the waxy locus of Zea mays, Mol. Gen. Genet. 203, 237-244 (1986). Promoters that express in the embryo, pericarp, and endosperm are disclosed in US applications Ser. Nos. 60/097,233 filed August 20, 1998 and 60/098,230 filed August 28, 1998. The disclosures each of these are incorporated herein by reference in their entirety.

Either heterologous or non-heterologous (i.e., endogenous) promoters can be employed to direct expression of the nucleic acids of the present invention. These promoters can also be used, for example, in expression cassettes to drive expression of antisense nucleic acids to reduce, increase, or alter concentration and/or composition of the proteins of the present invention in a desired tissue.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added can be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

An intron sequence can be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates. See for example Buchman and Berg, *Mol.*

Cell Biol. 8:4395-4405 (1988); Callis *et al.*, *Genes Dev.* 1:1183-1200 (1987). Use of maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. See generally, *The Maize Handbook*, Chapter 116, Freeling and Walbot, Eds., Springer, New York (1994).

The vector comprising the sequences from a polynucleotide of the present invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. Usually, the selectable marker gene will encode antibiotic or herbicide resistance. Suitable genes include those coding for resistance to the antibiotic spectinomycin or streptomycin (e.g., the *aada* gene), the streptomycin phosphotransferase (SPT) gene coding for streptomycin resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance.

Suitable genes coding for resistance to herbicides include those which act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance in particular the S4 and/or Hra mutations), those which act to inhibit action of glutamine synthase, such as phosphinothricin or basta (e.g., the *bar* gene), or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta and the ALS gene encodes resistance to the herbicide chlorsulfuron.

Typical vectors useful for expression of nucleic acids in higher plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers *et al.*, *Meth. In Enzymol.* 153:253-277 (1987). Exemplary *A. tumefaciens* vectors useful herein are plasmids pKYLX6 and pKYLX7 of Schardl *et al.*, *Gene* 61:1-11 (1987) and Berger *et al.*, *Proc. Natl. Acad. Sci. USA* 86:8402-8406 (1989). Another useful vector herein is plasmid pBI101.2 that is available from Clontech Laboratories, Inc. (Palo Alto, CA).

A variety of plant viruses that can be employed as vectors are known in the art and include cauliflower mosaic virus (CaMV), geminivirus, brome mosaic virus, and tobacco mosaic virus.

A polynucleotide of the present invention can be expressed in either sense or anti-sense orientation as desired. In plant cells, it has been shown that

antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, see, e.g., Sheehy *et al.*, *Proc. Nat'l. Acad. Sci. USA* 85: 8805-8809 (1988); and Hiatt *et al.*, U.S. Patent No. 4,801,340.

Another method of suppression is sense suppression. Introduction of nucleic acid configured in the sense orientation has been shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli *et al.*, *The Plant Cell* 2: 279-289 (1990) and U.S. Patent No. 5,034,323.

A method of down-regulation of the protein involves using PEST sequences that provide a target for degradation of the protein.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of plant genes. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. The design and use of target RNA-specific ribozymes is described in Haseloff *et al.*, *Nature* 334:585-591 (1988).

A variety of cross-linking agents, alkylating agents and radical generating species as pendant groups on polynucleotides of the present invention can be used to bind, label, detect, and/or cleave nucleic acids. For example, Vlassov, V. V., *et al.*, *Nucleic Acids Res* (1986) 14:4065-4076, describe covalent bonding of a single-stranded DNA fragment with alkylating derivatives of nucleotides complementary to target sequences. A report of similar work by the same group is that by Knorre, D. G., *et al.*, *Biochimie* (1985) 67:785-789. Iverson and Dervan also showed sequence-specific cleavage of single-stranded DNA mediated by incorporation of a modified nucleotide which was capable of activating cleavage (*J. Am. Chem. Soc.* (1987) 109:1241-1243). Meyer, R. B., *et al.*, *J. Am. Chem. Soc.* (1989) 111:8517-8519, effect covalent crosslinking to a target nucleotide using an alkylating agent complementary to the single-stranded target nucleotide sequence. A photoactivated crosslinking to single-stranded oligonucleotides mediated by psoralen was disclosed by Lee, B. L., *et al.*, *Biochemistry* (1988) 27:3197-3203. Use of crosslinking in triple-helix forming probes was also disclosed by Home, *et al.*, *J. Am. Chem. Soc.* (1990) 112:2435-2437. Use of N⁴, N⁴-ethanocytosine as an alkylating agent to crosslink to single-stranded oligonucleotides has also been described by Webb and Matteucci, *J. Am. Chem. Soc.* (1986) 108:2764-2765; *Nucleic Acids Res.* (1986) 14:7661-7674; Feteritz *et*

al., *J. Am. Chem. Soc.* 113:4000 (1991). Various compounds to bind, detect, label, and/or cleave nucleic acids are known in the art. See, for example, U.S. Patent Nos. 5,543,507; 5,672,593; 5,484,908; 5,256,648; and, 5,681,941.

PROTEINS

Proteins of the present invention include proteins derived from the native protein by deletion (so-called truncation), addition or substitution of one or more amino acids at one or more sites in the native protein. Such variants may result from, for example, genetic polymorphism or from human manipulation. Methods for such manipulations are generally known in the art.

For example, amino acid sequence variants of the polypeptide can be prepared by mutations in the cloned DNA sequence encoding the native protein of interest. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Walker and Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York); Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) *Methods Enzymol.* 154:367-382; Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, New York); U.S. Patent No. 4,873,192; and the references cited therein; herein incorporated by reference. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff *et al.* (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.), herein incorporated by reference. Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be preferred.

In constructing variants of the proteins of interest, modifications to the nucleotide sequences encoding the variants will be made such that variants continue to possess the desired activity. Obviously, any mutations made in the DNA encoding the variant protein must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure. See EP Patent Application Publication No. 75,444.

The isolated proteins of the present invention include a polypeptide comprising at least 23 contiguous amino acids encoded by any one of the nucleic acids of the present invention, or polypeptides which are conservatively modified variants thereof. The proteins of the present invention or variants thereof can

amino acids in length may be accomplished by attaching the C-terminal amino acid of the sequence to an insoluble support followed by sequential addition of the remaining amino acids in the sequence. Techniques for solid phase synthesis are described by Barany and Merrifield, *Solid-Phase Peptide Synthesis*, pp. 3-284 in *The Peptides: Analysis, Synthesis, Biology. Vol. 2: Special Methods in Peptide Synthesis, Part A.*; Merrifield, *et al.*, *J. Am. Chem. Soc.* 85:2149-2156 (1963), and Stewart *et al.*, *Solid Phase Peptide Synthesis, 2nd ed.*, Pierce Chem. Co., Rockford, Ill. (1984). Proteins of greater length may be synthesized by condensation of the amino and carboxy termini of shorter fragments. Methods of forming peptide bonds by activation of a carboxy terminal end (e.g., by the use of the coupling reagent N,N'-dicyclohexylcarbodiimide) is known to those of skill.

The proteins of this invention may be purified to substantial purity by standard techniques well known in the art, including detergent solubilization, selective precipitation with such substances as ammonium sulfate, column chromatography, immunopurification methods, and others. See, for instance, R. Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag: New York (1982); Deutscher, *Guide to Protein Purification*, Academic Press (1990). For example, antibodies may be raised to the proteins as described herein. Purification from *E. coli* can be achieved following procedures described in U.S. Patent No. 4,511,503. Detection of the expressed protein is achieved by methods known in the art and include, for example, radioimmunoassays, Western blotting techniques or immunoprecipitation.

The present invention further provides a method for modulating (i.e., increasing or decreasing) the concentration or composition of the polypeptides of the present invention in a plant or part thereof. Modulation of the polypeptides can be effected by increasing or decreasing the concentration and/or the composition of the polypeptides in a plant. The method comprises transforming a plant cell with an expression cassette comprising a polynucleotide of the present invention to obtain a transformed plant cell, growing the transformed plant cell under plant forming conditions, and expressing the polynucleotide in the plant for a time sufficient to modulate concentration and/or composition of the polypeptides in the plant or plant part.

In some embodiments, the content and/or composition of polypeptides of the present invention in a plant may be modulated by altering, *in vivo* or *in vitro*,

the promoter of a non-isolated gene of the present invention to up- or down-regulate gene expression. In some embodiments, the coding regions of native genes of the present invention can be altered via substitution, addition, insertion, or deletion to decrease activity of the encoded enzyme. See, e.g., Kmiec, U.S. Patent 5,565,350; Zarling *et al.*, PCT/US93/03868.

In some embodiments, an isolated nucleic acid (e.g., a vector) comprising a promoter sequence is transfected into a plant cell. Subsequently, a plant cell comprising the isolated nucleic acid is selected for by means known to those of skill in the art such as, but not limited to, Southern blot, DNA sequencing, or PCR analysis using primers specific to the promoter and to the nucleic acid and detecting amplicons produced therefrom. A plant or plant part altered or modified by the foregoing embodiments is grown under plant forming conditions for a time sufficient to modulate the concentration and/or composition of polypeptides of the present invention in the plant. Plant forming conditions are well known in the art.

In general, concentration of the polypeptides is increased or decreased by at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% relative to a native control plant, plant part, or cell lacking the aforementioned expression cassette. Modulation in the present invention may occur during and/or subsequent to growth of the plant to the desired stage of development.

Modulating nucleic acid expression temporally and/or in particular tissues can be controlled by employing the appropriate promoter operably linked to a polynucleotide of the present invention in, for example, sense or antisense orientation as discussed in greater detail above. Induction of expression of a polynucleotide of the present invention can also be controlled by exogenous administration of an effective amount of inducing compound. Inducible promoters and inducing compounds that activate expression from these promoters are well known in the art.

In preferred embodiments, the polypeptides of the present invention are modulated in monocots or dicots, preferably corn, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, millet, *Arabidopsis thaliana*, tomato, *Brassica* vegetables, peppers, potatoes, apples, spinach, or lettuce.

Means of detecting the proteins of the present invention are not critical aspects of the present invention. In a preferred embodiment, the proteins are detected and/or quantified using any of a number of well recognized

immunological binding assays (see, e.g., U.S. Patents 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also *Methods in Cell Biology*, Vol. 37: *Antibodies in Cell Biology*, Asai, Ed., Academic Press, Inc. New York (1993); *Basic and Clinical Immunology* 7th Edition, Stites & Terr, Eds. (1991). Moreover, the immunoassays of the present invention can be performed in any of several configurations, e.g., those reviewed in *Enzyme Immunoassay*, Maggio, Ed., CRC Press, Boca Raton, Florida (1980); Tijan, Practice and Theory of Enzyme Immunoassays, *Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers B.V., Amsterdam (1985); Harlow and Lane, *supra*; *Immunoassay: A Practical Guide*, Chan, Ed., Academic Press, Orlando, FL (1987); *Principles and Practice of Immunoassays*, Price and Newman Eds., Stockton Press, NY (1991); and *Non-isotopic Immunoassays*, Ngo, Ed., Plenum Press, NY (1988).

Typical methods for detecting proteins include Western blot (immunoblot) analysis, analytic biochemical methods such as electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, and the like, and various immunological methods such as fluid or gel precipitin reactions, immunodiffusion (single or double), immunoelectrophoresis, radioimmunoassays (RIAs), enzyme-linked immunosorbent assays (ELISAs), immunofluorescent assays, and the like.

Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (e.g., biotin) is covalently bound to the molecule. The ligand then binds to an anti-ligand (e.g., streptavidin) molecule that is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. A number of ligands and anti-ligands can be used. Where a ligand has a natural anti-ligand, for example, biotin, thyroxine, and cortisol, it can be used in conjunction with the labeled, naturally occurring anti-ligands. Alternatively, any haptenic or antigenic compound can be used in combination with an antibody.

The molecules can also be conjugated directly to signal generating compounds, *e.g.*, by conjugation with an enzyme or fluorophore. Enzymes of interest as labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases, or oxidoreductases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives,

dansyl, umbelliferone, etc. Chemiluminescent compounds include luciferin, and 2,3-dihydrophthalazinediones, e.g., luminol. For a review of various labeling or signal producing systems which may be used, see, U.S. Patent No. 4,391,904, which is incorporated herein by reference.

Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need be labeled and the presence of the target antibody is detected by simple visual inspection.

The proteins of the present invention can be used for identifying compounds that bind to (e.g., substrates), and/or increase or decrease (i.e., modulate) the enzymatic activity of, catalytically active polypeptides of the present invention. The method comprises contacting a polypeptide of the present invention with a compound whose ability to bind to or modulate enzyme activity is to be determined. The polypeptide employed will have at least 20%, preferably at least 30% or 40%, more preferably at least 50% or 60%, and most preferably at least 70% or 80% of the specific activity of the native, full-length polypeptide of the present invention (e.g., enzyme). Methods of measuring enzyme kinetics are well known in the art. See, e.g., Segel, *Biochemical Calculations*, 2nd ed., John Wiley and Sons, New York (1976).

Antibodies can be raised to a protein of the present invention, including individual, allelic, strain, or species variants, and fragments thereof, both in their naturally occurring (full-length) forms and in recombinant forms. Additionally, antibodies are raised to these proteins in either their native configurations or in non-native configurations. Anti-idiotypic antibodies can also be generated. Many methods of making antibodies are known to persons of skill.

In some instances, it is desirable to prepare monoclonal antibodies from various mammalian hosts, such as mice, rodents, primates, humans, etc. Description of techniques for preparing such monoclonal antibodies are found in, e.g., *Basic and Clinical Immunology*, 4th ed., Stites et al., Eds., Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane, *Supra*; Goding, *Monoclonal Antibodies: Principles and Practice*, 2nd ed.,

introduced heterologous nucleic acid. These seeds can be grown to produce plants that would produce the selected phenotype.

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are included in the invention, provided that these parts comprise cells comprising the isolated nucleic acid of the present invention. Progeny and variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced nucleic acid sequences.

Transgenic plants expressing a selectable marker can be screened for transmission of the nucleic acid of the present invention by, for example, standard immunoblot and DNA detection techniques. Transgenic lines are also typically evaluated on levels of expression of the heterologous nucleic acid. Expression at the RNA level can be determined initially to identify and quantitate expression-positive plants. Standard techniques for RNA analysis can be employed and include PCR amplification assays using oligonucleotide primers designed to amplify only the heterologous RNA templates and solution hybridization assays using heterologous nucleic acid-specific probes. The RNA-positive plants can then be analyzed for protein expression by Western immunoblot analysis using the specifically reactive antibodies of the present invention. In addition, *in situ* hybridization and immunocytochemistry according to standard protocols can be done using heterologous nucleic acid specific polynucleotide probes and antibodies, respectively, to localize sites of expression within transgenic tissue. Generally, a number of transgenic lines are usually screened for the incorporated nucleic acid to identify and select plants with the most appropriate expression profiles.

A preferred embodiment is a transgenic plant that is homozygous for the added heterologous nucleic acid; i.e., a transgenic plant that contains two added nucleic acid sequences, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) a heterozygous transgenic plant that contains a single added heterologous nucleic acid, germinating some of the seed produced and analyzing the resulting plants produced for altered expression of a polynucleotide of the present invention relative to a control plant (i.e., native, non-transgenic). Back-

crossing to a parental plant and out-crossing with a non- transgenic plant are also contemplated.

The present invention provides a method of genotyping a plant comprising a polynucleotide of the present invention. Genotyping provides a means of distinguishing homologs of a chromosome pair and can be used to differentiate segregants in a plant population. Molecular marker methods can be used for phylogenetic studies, characterizing genetic relationships among crop varieties, identifying crosses or somatic hybrids, localizing chromosomal segments affecting monogenic traits, map based cloning, and the study of quantitative inheritance. See, e.g., *Plant Molecular Biology: A Laboratory Manual*, Chapter 7, Clark, Ed., Springer-Verlag, Berlin (1997). For molecular marker methods, see generally, *The DNA Revolution* by Andrew H. Paterson 1996 (Chapter 2) in: *Genome Mapping in Plants* (ed. Andrew H. Paterson) by Academic Press/R. G. Landis Company, Austin, Texas, pp.7-21.

Plants that can be used in the method of the invention include monocotyledonous and dicotyledonous plants. Preferred plants include corn, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, millet, *Arabidopsis thaliana*, tomato, *Brassica* vegetables, peppers, potatoes, apples, spinach, or lettuce.

The present invention will be further described by reference to the following detailed examples. It is understood, however, that there are many extensions, variations, and modifications on the basic theme of the present invention beyond that shown in the examples and description, which are within the spirit and scope

of the present invention. All publications, patents, and patent applications cited herein are hereby incorporated by reference.

EXAMPLES

Identification of a phytyl/prenyltransferase involved in biosynthesis of tocopherols in *Synechocystis* PCC 6803 and *Arabidopsis thaliana*.

SubB3
00324-04200

PCC 6803 was used as a tool for identification of genes encoding enzymes involved in biosynthesis of tocopherols. *Synechocystis* is a cyanobacterium capable of tocopherol biosynthesis. The entire genome of this photosynthetic organism has been recently sequenced (Kaneko et al., 1996) and the data is available on a public searchable database, called CyanoBase (<http://www.kazusa.or.jp/cyano/cyano.html>). Using CyanoBase, we have identified an open reading frame (SLR1736) encoding a phytyl/prenyltransferase involved in the biosynthesis of 2-methyl-6-phytylplastoquinol, one of the tocopherol precursors. This open reading frame was identified based on similarity with the phytyl/prenyltransferase SLR0056, a phytyl/prenyltransferase involved in the biosynthesis of chlorophyll in *Synechocystis* PCC 6803. SLR0056 exhibits a high homology with the previously identified chlorophyllide/phytyl/prenyltransferases from many cyanobacteria and *A. thaliana* (Lopez et al., 1996), suggesting that this enzyme is also involved in chlorophyll synthesis.

SLR1736 is similar, but not highly homologous to the SLR0056 open reading frame. However, the putative prenyl-binding domain is highly conserved in SLR1736, making it a good candidate for the tocopherol phytyl/prenyltransferase. Using the SLR1736 translated sequence as a query in the blast search, a genomic clone on chromosome II was identified in the *A. thaliana* database (Stanford Genomic Resources). This genomic clone was used to isolate an *Arabidopsis* cDNA clone. The F19F24 genomic clone and *Arabidopsis* cDNA are highly homologous to the SLR1736 protein sequence. The prenyl-binding domain is also conserved in the F19F24 and *Arabidopsis* cDNA. In addition, the amino terminal deduced amino acid sequence of the *Arabidopsis* gene and cDNA exhibits the traits of chloroplast-targeting sequences. Tocopherol biosynthesis has been shown to take place in chloroplast envelopes (Soll et al., 1980; Soll, 1987). We believe that the *Arabidopsis* F19F24 gene and homologous

cDNA represent the orthologous phytyl/prenyltransferase that attaches phytyldiphosphate (phytyl-PP) and/or geranylgeranyl pyrophosphate (GGPP) to homogentisic acid in tocopherol synthesis in *A. thaliana*. Additionally, a 1.2 kb corn EST, chste82, that is highly homologous to SLR1736 has been also identified in a blast search.

To demonstrate that SLR1736 might be involved in tocopherol biosynthesis in *Synechocystis*, this gene was disrupted by insertion of the kanamycin expression cassette. The method of gene disruption by gene replacement technique has been previously described (Williams, 1988). The resulting mutant was named Δ SLR1736. Before analyses, the mutant was sub-cultured at least 6 times by single colony section on kanamycin to select for the colonies containing only copies of the SLR1736 gene disrupted with the kanamycin resistance gene. The absence of wild type SLR1736 genes was confirmed by PCR. The lack of tocopherols in the mutant was shown by HPLC separation of lipid extracts from wild type and mutant *Synechocystis* on a normal-phase column using fluorescent detection (FLD). Levels of phylloquinone (vitamin K1) and plastoquinone seem to be unaffected in this mutant. This suggests that there are at least two separate prenyltransferase activities for tocopherol and plastoquinone synthesis in *Synechocystis* and we may be able to manipulate carbon flow through the pathway by altering gene expression of either. Phytylation/prenylation of homogentisic acid is the branch-point in tocopherol and plastoquinone synthesis, and therefore, most likely an important regulatory point of the pathway. As well as the prenylation activities, availability of different prenyl tails may also be crucial for the regulation of carbon flow through the pathway. This will become significant for manipulating tocopherol levels in higher plants.

Amplification of the SLR1736 open reading frame from *Synechocystis*

Chromosomal DNA from wild type *Synechocystis* PCC 6803 was isolated according to Williams (Methods in Enzymology (1987) 167: 766-778). The following primers were designed using Mac Vector computer program to amplify a 1.022 kb fragment containing the SLR1736 open reading frame:

SLR1736F: 5'-TATTCATATGGCAACTATCCAAGCTTTTGTG-3'

SLR1736R: 5'-GGATCCTAATTGAAGAAGATACTAAATAGTTC-3'

SubB4
NdeI and BamHI sites were added to the primers to facilitate sub-cloning for expression purposes. ATG in the SLR1736F primer is the start codon for the SLR1736 open reading frame published in the CyanoBase Web-site. Taq polymerase (Gibco BRL) was used for gene disruption purposes and later Vent polymerase (NEB) was used for expression purposes following the manufacturer's recommendations. The following cycles were performed:

For Taq polymerase amplification:

95 °C/5 minutes (1 cycle)

95 °C/45 seconds, 45 °C/45 seconds, 68 °C/45 seconds (5 cycles)

95 °C/45 seconds, 52 °C/45 seconds, 72 °C/45 seconds (30 cycles)

72 °C/10 minutes

The same thermocycler conditions were used to amplify SLR1736 with Vent polymerase except that elongation times were extended to 2 minutes.

Sub-cloning the SRL1736 PCR product

Plasmid pBluescript KS II (Stratagene) was digested with EcoRV (NEB) according to manufacturer's protocols. Both linearized pBluescript and the amplified SLR1736 open reading frame were separated by 0.9 % agarose TBE gel electrophoresis. The bands were excised and purified from the gel using a JetSorb DNA purification kit (PGC Scientifics). The purified fragment was sub-cloned into the EcoRV site of pBluescript KS II in a blunt-end ligation reaction. A 10 µl ligation reaction contained 20 mM Tris-HCl (pH 7.6), 5 mM MgCl₂, 5 mM DTT, 50 µg/ml BSA, 0.5 mM rATP, 15 % PEG, and 1 U of T4 DNA ligase (Gibco BRL). Ligation was carried out at room temperature for 4 hours. One half of the reaction mixture was used to transform competent *E. coli* DH5α cells. Transformants were then selected on LB plates containing 100 mg/L of ampicillin. X-gal and IPTG were used for blue/white selection. White ampicillin resistant colonies were then selected, grown in liquid LB/ampicillin media, and plasmids were purified. The resulting plasmid was designated as KS-1736 and the nature and the orientation of the 1736 insert was determined by restriction digestion and sequencing (ABI Prism 310 Genetic Analyser). Clone #5, in which SLR1736 was in a reverse orientation to the Lac promoter of the vector, was selected for further manipulations.

The SLR1736 replacement construct

Transformation followed by homologous recombination is feasible in *Synechocystis* (Williams, 1988). A gene of interest, in our case SLR1736, can be easily disrupted by inserting an antibiotic resistance gene into the coding region. Such a disruption construct can be transformed into *Synechocystis*. pBluescript KS or any other vector capable of replication in *E. coli* can be used as a vector. These vectors cannot be replicated by DNA replication machinery of *Synechocystis* so that the cells are forced to keep the resistance gene by other means when kept on the antibiotic selection. In *Synechocystis*, the wild type copies of the target gene are replaced with the copies of this gene disrupted with the antibiotic resistance cassette by homologous recombination. Since this cyanobacterium contains multiple copies of its genome, it is necessary to streak selected resistant colonies on the selection media several times. This should ensure replacement of the wild type copies of the gene with the disrupted ones (Williams, 1988).

The kanamycin resistance gene from the transposon Tn903 encoding aminoglycoside 3'-phosphotransferase was used to disrupt the wild type SLR1736 gene. Plasmid pUC4K (Pharmacia) was cut with EcoRI to release the kanamycin resistance expression cassette. Since SLR1736 has a unique MfeI site about 200 bp from the beginning of the gene, plasmid KS-1736 #5 was digested with MfeI (NEB). MfeI leaves 5'-cohesive ends compatible with EcoRI so that no other molecular manipulations are necessary. The two DNA fragments were purified from agarose gels as described above and ligated using T4 DNA ligase (Gibco BRL) as recommended by the manufacturer. Competent *E. coli* DH5 α cells were transformed with the ligation reaction and transformants selected on LB plates containing 50 mg of kanamycin per liter of media. Plasmids were purified and subjected to restriction analysis. Two plasmids having opposite orientation of the kanamycin cassette were chosen for *Synechocystis* transformation. The two constructs were designated as KS Δ 1736-KAN-F and B, respectively, to indicate the orientation of the kanamycin resistance gene in respect to the SLR1736 gene.

Transformation of *Synechocystis* PCC 6803 with KS Δ 1736-KAN-F and B, respectively, was carried out as described by Williams (1988). Transformants were selected on BG-11 plates containing 15 mM glucose and 5 mg of kanamycin

per liter of medium. Two independent colonies from each transformation were then sub-cultured once a week for several weeks on BG-11 plates containing 15 mM glucose and 15 mg/L kanamycin before being analyzed. The cells were grown under continuous light at 30°C. The resulting clones used for further analyses were designated Δ SLR1736 F-1, F-2, B-1, and B-2.

Confirmation of the SLR1736 gene disruption by PCR

Chromosomal DNA from wild type and Δ SLR1736 mutant *Synechocystis* PCC 6803 was isolated from a few colonies according to Cai and Wolk (1990) with minor modifications as follows: Cells were resuspended in 200 μ l of 50 mM Tris.HCl and 10 mM EDTA solution of pH 7.5. The cells were then transferred to a 2 ml screw-cup tube. 10 μ l of 20% SDS, 200 μ l of phenol:chloroform (1:1), and white sand were added. The samples were mixed well by vortexing for 1 minute and then they were placed on ice for another minute. This step was repeated twice. The mixture was centrifuged at 14,000 rpm for 5 minutes to separate organic and aqueous layers. The top aqueous phase was then extracted twice with an equal volume of chloroform and precipitated with a quarter of volume of 3M potassium acetate (pH 4.8) and two volumes of 96% of ethanol. After an hour incubation at -20°C and a ten-minute centrifugation at 14,000 rpm, genomic DNA was washed once with 80% of ethanol, dried in a speedvac for 2-3 minutes, and resuspended in 20 μ l of water. DNA was diluted 1:10 with water and used as a template in PCR reactions. PCR was performed as described above using Taq polymerase (Gibco BRL). Insertion of the kanamycin cassette into the SLR1736 open reading frame was clearly demonstrated.

HPLC analyses of the lipid extracts from wild type and mutant *Synechocystis*.

Tocopherol analysis

About 30 mg of wild type and Δ SLR1736 F and B mutant cells grown on solid plates as described above were harvested and resuspended in 450 μ l of methanol: chloroform (2:1) containing 1 mg/ml of butylated hydroxytoluene (BHT) to prevent oxidation of tocopherols. 200 ng of tocol was added as an internal standard. The cells were homogenized using mini-pestals followed by addition of 150 μ l of chloroform and 300 μ l of water to the mixture. After centrifugation (5

minutes at 14,000 rpm), the lower organic phase was transferred to a clean microfuge tube and dried in a speedvac. Lipids were resuspended in 80 μ l of hexane containing 1 mg/ml of BHT. 40 μ l of the lipid extract was subjected to HPLC (Hewlett-Packard 1100 Series HPLC system with a fluorescence detector) using a normal phase column (Lichrosorb Si60A 4.6 X 250 mm) equilibrated at 42°C. A 20-minute linear gradient of 8 % to 18 % di-isopropyl ether in hexane was used to separate different types of tocopherols. After excitation at a wavelength of 290 nm, tocopherols were detected by their fluorescence at 325 nm.

Wild type *Synechocystis* accumulates predominantly α -tocopherol. The Δ SLR1736 disruption mutants lack all tocopherols and this effect is independent of the kanamycin cassette orientation. These results indicate that SLR1736 is involved in tocopherol biosynthesis and acts upstream of the methyltransferases. Disruption of the methyltransferase genes SLL0418 (2-methyl-6-phytylplastoquinol methyltransferase) and SLR0089 (γ -tocopherol methyltransferase) which have been recently cloned from *Synechocystis* leads to the accumulation of β - and γ -tocopherols, respectively (Shintani, D., personal communication; Shintani and DellaPenna, 1998). The only two possible remaining enzymes are the cyclase and prenyltransferase. Since SLR1736 exhibits a similarity to known prenyltransferases, we believe this enzyme represents a prenyltransferase. More conclusive proof than the one based on similarity would be given by in vitro prenyltransferase assays and feeding studies of wild type and Δ SLR1736 mutant *Synechocystis* with 14 C uniformly labeled tyrosine.

Phylloquinone and plastoquinone analysis

Formation of homogentisic acid, the first step of the pathway, is common for both tocopherols and plastoquinone in photosynthetic organisms. To answer the question if the tocopherol prenyltransferase is also involved in plastoquinone biosynthesis and how carbon flow is affected in the plastoquinone part of the pathway, we analyzed lipid extracts from wild type and the mutant cells. On the other hand, the phytyl tail is a part of vitamin K1 (phylloquinone) molecule. To estimate effects of the SLR1736 gene disruption on the phylloquinone biosynthesis in *Synechocystis*, we also performed vitamin K1 analysis.

About 30 mg of wild type and Δ SLR1736 F and B mutant were harvested and resuspended in 450 μ l of methanol: chloroform (2:1). The cells were homogenized using mini-pestals followed by addition of 150 μ l of chloroform and 300 μ l of water to the mixture. After centrifugation (5 minutes at 14,000 rpm), the lower organic phase was transferred to a clean microfuge tube, dried in a speedvac, dissolved in 30 μ l of ethyl acetate, and oxidized with silver oxide for a half an hour. The entire extract was loaded on a TLC plate (Silica, 60A) which was developed in 20% diethyl ether in petroleum ether and dried. The plate was sprayed with leucomethylene blue (Crane & Barr, 1971) to visualize any changes in quinone composition. No differences between the wild type and mutant quinone profiles were observed. To prepare leucomethylene blue, 50 mg of methylene blue and 0.5 g of zinc dust were mixed in 5 ml of water. The mixture was acidified with a few drops of concentrated sulfuric acid and left to react for about 10 minutes before use.

To quantify possible changes in quinone content in wild type and mutant *Synechocystis*, HPLC analyses of lipid extracts containing plastoquinone-8 as an internal standard were performed. Lipids were extracted as described above except that 1 μ g of plastoquinone-8 was added in the beginning of extraction. Plastoquinone-8 and plastoquinone-9 standards were isolated and purified from *Iris holandica* bulbs (Hutson & Therfall, 1980) and their concentrations were determined using the molar absorption coefficient of plastoquinone-9 at 254 nm, 17.94 mM⁻¹ cm⁻¹. These quinones have similar properties and they can be easily separated by the HPLC method described below. Therefore, plastoquinone-8 is an excellent internal standard.

After extraction, quinones were resuspended in 80 μ l of HPLC grade ethyl acetate. 40 μ l of the lipid extract was subjected to HPLC (Hewlett-Packard 1100 Series HPLC system) using a C-18 reverse phase column (Spherisorb, 4.6 X 250 mm). The following conditions were utilized to separate different quinones:

Time (min.):	Reagent Alcohol (%):	Water (%):
0	90	10
1	99	1
10	99	1
11	100	0
16	100	0
17	90	10
35	90	10

The flow of the solvents was 0.8 ml/min and the separation was performed at room temperature. Quinones were detected by their absorbance at 250 and 275 nm using a diode array detector and the identity of phylloquinone and plastoquinone-9 was confirmed by comparison with their previously published spectra (Crane & Barr, 1971). No differences in vitamin K1 and plastoquinone-9 compositions were observed between wild type and the Δ SLR1736 disruption mutant. This indicates that the SLR1736 gene product is involved only in tocopherol biosynthesis.

Cloning phytyl transferase from *A. thaliana* involved in tocopherol biosynthesis

A developing seed-specific cDNA library from *A. thaliana* (lambda-ZAP type, provided by John Ohlrogge at the Michigan State University) was screened using a PCR product from wild type *A. thaliana* genomic DNA (Ler ecotype) which exhibits a high degree of homology with the *Synechocystis* phytyl transferase. Primers AT1736F (5'-TTGTTTTTCAGGCTGTTGTTGCAGCTCTC-3') and AT1736B (5'-CGTTTCTGACCCAGAGTTACAGAGAATG-3') were used to amplify about 1kb fragment corresponding to 60238 – 61229 bp region of the BAC clone F19F24 (*A. thaliana* database at Stanford). The following program was used to amplify this fragment with Vent DNA polymerase (New England Biolabs):

95 °C/5 minutes (1 cycle)

95 °C/45 seconds; 50 °C/45 seconds, 72 °C/1 minute (30 cycles)

72 °C/10 minutes (1 cycle)

The PCR product was then sub-cloned into *EcoRV* site of pBluescript KS (Stratagene) as in the case of the cyanobacterial phytyl transferase presented above and sequenced from both ends using T3 and T7 primers (Stratagene) to

Sub B6

ensure the identity of the sub-cloned fragment. A 300 bp fragment of the insert (5'-end) was released with *EcoRI* from the vector and used as a radioactively-labeled probe to obtain full-length clones. About 2.5 million plaques of the seed-specific library were screened using standard procedures (Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning*. 2nd edition, Cold Spring Harbor Laboratory Press). 16 positive non-purified plaques were chosen for PCR analysis using T3 and AT1736T7c (5'-GACATATTTTGCAGTCTGCC-3) which is an internal primer for the phytyl transferase. Clones #1, 3, 5, 8, 11, 12, and 14 were selected for further purification and single clone excision, performed according to manufacturer (Stratagene), to obtain individual clones in pBluescript SK plasmids. Each clone was sequenced from each end using T3 and T7 primer. The longest clone, #11 – about 1.6 kb, was chosen for complete sequencing which is in progress now. All clones were aligned to the genomic clone F19F24 from *A. thaliana* to confirm their identity, identify introns and find possible sequencing mistakes in the genomic sequence. We believe that ATG codon (59220 bp on F19F24) is the start codon of the phytyl transferase involved in tocopherol synthesis in *A. thaliana*. Starting from this methionine, the first 36 amino acids represent the chloroplast thylakoid membrane-targeting sequence (PSORT program, <http://psort.nibb.ac.jp:8800/>).

Confirmation of prenyltransferase nature of SLR1736

To confirm the prenyltransferase nature of SLR1736, the intact gene will be expressed in *E. coli* because this bacterium lacks any enzymatic activity connected to tocopherol biosynthesis. Therefore, SLR1736 activity will be shown by an in vitro phytyl/prenyltransferase assay using protein extracts from *E. coli* expressing SLR1736 or by reconstruction of multiple steps of the pathway in *E. coli*. ¹⁴C uniformly labeled p-hydroxyphenyl pyruvate and phytyl-PP, or other prenyl diphosphates will be used as substrates. p-hydroxyphenyl pyruvate dioxygenase catalyses conversion of p-hydroxyphenylpyruvic acid to homogentisic acid, the immediate substrate for the tocopherol and plastoquinone prenyltransferase(s). Therefore, *A. thaliana* p-hydroxyphenylpyruvic acid dioxygenase (Norris *et al.*, 1998) expressed in *E. coli* along with the prenyltransferase will be present in the reactions to couple the two enzymatic steps. To further show that SLR1736 is a prenyltransferase, Δ SLR1736 and wild

type *Synechocystis* will be grown in the presence of ^{14}C uniformly labeled L-tyrosine to trace prenylated products by using TLC and autoradiography.

The SLR1736 open reading frame will be also expressed in *E. coli* in the presence of p-hydroxyphenylpyruvic acid dioxygenase (Norris *et al.*, 1998), *Adonis paleastina* geranylgeranyl diphosphate synthase (gift from F. Cunningham), and geranylgeranyl hydrogenase from *Synechocystis* (SLL1091, Addlesee *et al.*, 1996; Keller *et al.*, 1998) to reconstitute the phytyl pyrophosphate pathway since *E. coli* does not possess any of these enzymatic activities. Lipids will be extracted and subjected to HPLC analysis of quinones as described above. 2-methyl-6-phytylplastoquinone is stable and should be present in *E. coli* lipid extracts.

SLR1736 homologue from *A. thaliana* (AT1736)

To investigate the role of the plant homologue of SLR1736, the intact full length cDNA from *Arabidopsis thaliana* (AT1736) and corn chste82 EST will be expressed in the sense and antisense orientation from the constitutive CaMV 35S or seed-specific (Seffens *et al.*, 1990) promoters, respectively, in *A. thaliana*. Visible phenotype(s) will be observed and lipids from the transgenic plants will be extracted and subjected to HPLC/FLD analyses to detect changes in tocopherol content and composition in green tissues and seeds. Plastoquinone and phyloquinone levels will also be analyzed as described above. It is possible that phytyl-PP is limiting for the prenyltransferase activity. Consequently, we may want to overexpress geranylgeranyl pyrophosphate synthase and GGDP dehydrogenase to elevate phytyl-PP levels in *A. thaliana*. Columbia ecotype *Arabidopsis* plants will be transformed with these overexpression constructs separately and homozygous transformants will be crossed to obtain plants containing all three constructs.

The *in vitro* prenyltransferase assay will be performed with AT1736 expressed in *E. coli* as described above for SLR1736. Prenyl tail-specificity studies will be also carried out with this enzyme, using various tails such as GGDP, phytyl-PP, and solanyl-PP. As in the case of SLR1736 from *Synechocystis*, it is important to determine if there are one or two prenyltransferases for tocopherol and plastoquinone biosynthesis in higher plants.

Construction of p0018 *Maize* cDNA libraries

Total RNA Isolation

Total RNA was isolated from p0018 library corn tissues with TRIzol Reagent (Life Technology Inc. Gaithersburg, MD) using a modification of the guanidine isothiocyanate/acid-phenol procedure described by Chomczynski and Sacchi (Chomczynski, P., and Sacchi, N. Anal. Biochem. 162, 156 (1987)). In brief, plant tissue samples were pulverized in liquid nitrogen before the addition of the TRIzol Reagent, and then were further homogenized with a mortar and pestle. Addition of chloroform followed by centrifugation was conducted for separation of an aqueous phase and an organic phase. The total RNA was recovered by precipitation with isopropyl alcohol from the aqueous phase.

Poly(A)+ RNA Isolation

The selection of poly(A)+ RNA from total RNA was performed using PolyATact system (Promega Corporation. Madison, WI). In brief, biotinylated oligo(dT) primers were used to hybridize to the 3' poly(A) tails on mRNA. The hybrids were captured using streptavidin coupled to paramagnetic particles and a magnetic separation stand. The mRNA was washed at high stringent condition and eluted by RNase-free deionized water.

cDNA Library Construction

cDNA synthesis was performed and unidirectional cDNA libraries were constructed using the SuperScript Plasmid System (Life Technology Inc. Gaithersburg, MD). The first strand of cDNA was synthesized by priming an oligo(dT) primer containing a Not I site. The reaction was catalyzed by SuperScript Reverse Transcriptase II at 45°C. The second strand of cDNA was labeled with alpha-32P-dCTP and a portion of the reaction was analyzed by agarose gel electrophoresis to determine cDNA sizes. cDNA molecules smaller than 500 base pairs and unligated adapters were removed by Sephacryl-S400 chromatography. The selected cDNA molecules were ligated into pSPORT1 vector in between of Not I and Sal I sites.

Sequencing of Maize cDNA and Library Subtraction

Sequencing Template Preparation

Individual colonies were picked and DNA was prepared either by PCR with M13 forward primers and M13 reverse primers, or by plasmid isolation. All the cDNA clones were sequenced using M13 reverse primers.

Q-bot Subtraction Procedure

cDNA libraries subjected to the subtraction procedure are plated out on 22 x 22 cm² agar plate at density of about 3,000 colonies per plate. The plates are incubated in a 37°C incubator for 12-24 hours. Colonies are picked into 384-well plates by a robot colony picker, Q-bot (GENETIX Limited). These plates are incubated overnight at 37°C.

Once sufficient colonies are picked, they were pinned onto 22 x 22 cm² nylon membranes using Q-bot. Each membrane contained 9,216 colonies or 36,864 colonies. These membranes are placed onto agar plate with appropriate antibiotic. The plates are incubated at 37°C for overnight.

After colonies are recovered on the second day, these filters are placed on filter paper prewetted with denaturing solution for four minutes, then are incubated on top of a boiling water bath for additional four minutes. The filters are then placed on filter paper prewetted with neutralizing solution for four minutes. After excess solution is removed by placing the filters on dry filter papers for one minute, the colony side of the filters are place into Proteinase K solution, incubated at 37°C for 40-50 minutes. The filters are placed on dry filter papers to dry overnight. DNA is then cross-linked to nylon membrane by UV light treatment.

Colony hybridization is conducted as described by Sambrook, J., Fritsch, E.F. and Maniatis, T., (in Molecular Cloning: A laboratory Manual, 2nd Edition). The following probes were used in colony hybridization:

1. First strand cDNA from the same tissue as the library was made from to remove the most redundant clones.
2. 48-192 most redundant cDNA clones from the same library based on previous sequencing data.
3. 192 most redundant cDNA clones in the entire corn sequence database.
4. A Sal-A20 oligo nucleotide: TCG ACC CAC GCG TCC GAA AAA AAA AAA AAA AAA AAA, removes clones containing a poly A tail but no cDNA.

5. cDNA clones derived from rRNA.

The image of the autoradiography is scanned into computer and the signal intensity and cold colony addresses of each colony is analyzed. Re-arraying of cold-colonies from 384 well plates to 96 well plates is conducted using Q-bot.

Identification of Gene from a Computer Homology Search

Gene identities were determined by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., *et al.*, (1993) J. Mol. Biol. 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches under default parameters for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. Nature Genetics 3:266-272 (1993)) provided by the NCBI. In some cases, the sequencing data from two or more clones containing overlapping segments of DNA were used to construct contiguous DNA sequences.

Construction of Additional *Maize, Rice, Soybean and Wheat* cDNA libraries

Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from maize, rice, soybean and wheat tissues were prepared. The characteristics of these libraries are described in Table 1.

TABLE 1		
Library Designation	Library Description	Clone
cco1n	Corn Cob of 67 Day Old Plants Grown in Green House*	cco1n.pk087.117
p0018	Seedling after 10 day drought (T001), heat shocked for 24 hrs (T002), recovery at normal growth condition for 8 hrs, 16 hrs, 24hrs	p0018.chste82r:fis
p0108	PR leaves + C.carbonium, screened 1 Pool of PR+C. carbonium tox-3h; PR+C. carbonium tox-6h; PR+C. carbonium tox-24h; PR+C. carbonium tox-48hr; and PR+C. carbonium tox-7 7 days	p0108.cjrmc89r:fis
rca1n	Rice (<i>Oryza sativa</i> L., Nipponbare) callus normalized.	rca1n.pk025.c4
rl0n	Rice 15 Day Old Leaf*	rl0n.pk0066.e2:fis
scr1c	Soybean (<i>Glycine max</i> L., 2872) Embryogenic suspension culture subjected to 4 vacuum cycles and collected 12 hrs later (control scb1c).	scr1c.pk005.l2
sgc7c	Soybean (<i>Glycine max</i> L., Wye) germanating cotyledon (yellow and wilting; 18-30 DAG).	sgc7c.pk001.h22
src2c	Soybean (<i>Glycine max</i> L., 437654) 8 day old root inoculated with eggs of cyst Nematode (Race 1) for 4 days.	src2c.pk020.d5:fis
wdk2c	Wheat Developing Kernel, 7 Days After Anthesis.	wdk2c.pk012.f2
wlm0	Wheat Seedlings 0 Hour After Inoculation With <i>Erysiphe graminis f. sp tritici</i>	wlm0.pk0011.c7

*These libraries were normalized essentially as described in U.S. Pat. No. 5,482,845, incorporated herein by reference.

In general, cDNA libraries may be prepared by the method described above or by any one of many other methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences.

Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or “ESTs”; see Adams et al., (1991) *Science* 252:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

Characterization of cDNA Clones Encoding Phytyl/prenyltransferase.

cDNA Clones were identified by computer homology search as described above. The BLASTP and BLASTN searches using the sequences from clones listed in Table 1 revealed similarity to certain polypeptides as shown in Table 2. The “/blast/data/2.0/2/nr” database was searched. GAP results showing % identity to synechocystis and arabidopsis are also shown. Table 2 shows the BLAST results for individual complete gene sequences (“CGS”).

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted March 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TABLE 2

Top BLAST Results for Sequences Encoding Polypeptides Homologous to Phytyl/prenyltransferase and GAP % Identity to Synechocystis and Arabidopsis

Clone	Status	Protein Sequence with Significant Alignment gi# (accession #) Organism; % Blast Identity	GAP % Identity Clone to D90909	GAP % Identity Clone to AC003673
SEQ ID 12 - Contig of: cco1n.pk087.117 and cen3n.pk0012.h6	CGS	1652856 (D90909) Synechocystis; 36% 3004556 (AC003673) Arabidopsis; 32%	36.80%	50.27%
SEQ ID 4 - p0018.chste82r: fis	CGS	1652856 (D90909) Synechocystis; 43% 3004556 (AC003673) Arabidopsis; 47%	43.58%	70.67%
SEQ ID 14 - p0108.cjrmc89r: fis	CGS	1652856 (D90909) Synechocystis; 36% 6015890 (Y18930) Sulfolobus; 30% 5103549 (AP000058) Aeropyrum; 32% 3004556 (AC003673) Arabidopsis; 26%	37.54%	30.30%
SEQ ID 16 - rca1n.pk025.c4	CGS	1652856 (D90909) Synechocystis; 45% 3004556 (AC003673) Arabidopsis; 43%	45.27%	70.67%
SEQ ID 18 - r10n.pk0066.e2: fis	CGS	1652856 (D90909) Synechocystis; 35% 5103549 (AP000058) Aeropyrum; 29% 6015890 (Y18930) Sulfolobus; 28% 3004556 (AC003673) Arabidopsis; 25%	35.59%	31.16%
SEQ ID 20 - scr1c.pk005.12	CGS	1652856 (D90909) Synechocystis; 37% 6015890 (Y18930) Sulfolobus; 28% 3004556 (AC003673) Arabidopsis; 25%	36.95%	33.33%
SEQ ID 22 - Contig of: sgc7c.pk001.h22	CGS	1652856 (D90909) Synechocystis; 44% 3004556 (AC003673) Arabidopsis; 83%	44.90%	75.48%
SEQ ID 24 - src2c.pk020.d5: fis	CGS	3004556 (AC003673) Arabidopsis; 39% 1652856 (D90909) Synechocystis; 29%	30.51%	52.62%
SEQ ID 26 - Contig of: wdk2c.pk012.f2	Partial Gene Seq	3004556 (AC003673) Arabidopsis; 45% 1652856 (D90909) Synechocystis; 37%	37.50%	43.75%
SEQ ID 28- Contig of: wlm0.pk0011.c7	CGS	1652856 (D90909) Synechocystis; 36% 3004556 (AC003673) Arabidopsis; 27%	37.54%	30.30%

Sequence alignments and BLAST sequence identities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of a phytyl/prenyltransferase.

A BLASTN search of "/blast/data/2.0/3/est" using the sequences from clones listed in Table 1 showed homology (E score > 140) to the following sequences on Table 3 as indicated by Genbank Accession number.

TABLE 3 - GAP Search Result

Number	Species	Shows Homology With Seq ID No.
C25006	Rice	17
C74444	Rice	15
AA750728	Rice	15
AA749638	Rice	15
AU029707	Rice	15
AI612332	Corn	13
AI711952	Corn	13
AI795680	Corn	13
AI897027	Tomato	21
AI938270	Soybean	21
AI938569	Soybean	21
		23
AI948381	Corn	13
AW052841	Corn	13
AW054141	Corn	11
AW066179	Corn	11
AW146615	Corn	13, 17
AW202246	Soybean	19
AI444024	Soybean	19
AI442111	Soybean	19
AW132909	Soybean	23
		21
AI748688	Soybean	21
AI939002	Soybean	19
AW306617	Soybean	21
AW433064	Soybean	21
AW563431	Sorghum	17

In sequencing clone containing SEQ ID NO: 11, an extra nucleotide at nt 631 was observed. In addition, possible frameshifts at nt 107-140 were located that may interrupt homology to the Synechocystis hypothetical protein # 1652856. The extra nucleotide at nt 631 was deleted from the sequence listing provided with this application, and sequence identity was determined without considering the extra nucleotide. The extra nucleotide is likely an artifact occurring during the isolation and sequencing of the cDNA clones.

Clones p0108.cjrmc89r.fis and r10n.pk0066.e2.fis each contain substantially complete gene sequence, with the exception of a few N-terminal amino acids on each.

Clone wdk2c.pk012.f2 has an apparent intron from nt 322 to 426, as determined by GT/AG intron borders, that interrupt homology to p0018.chste82r.fis. The sequence listing provides both the nucleotide sequence of the clone with the intron (SEQ ID NO: 29) and without the intron (SEQ ID NO: 25). Amino acid sequence identity in SEQ ID NO: 26 was determined after removal of the intron.

The amino acid sequence of clone wlm0.pk0011.c7 covers the entire phytyl/prenyltransferase and contains a putative transit peptide sequence.

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb Sall-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-Sall fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli*

XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 µL of a 2.5 M solution) and spermidine free base (20 µL of a 1.0 M solution) are added to the particles. The

suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μ L of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 μ L of ethanol. An aliquot (5 μ L) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology* 8:833-839).

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol.*

Chem. 261:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with fluorescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) 327:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the

T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 μL of a 60 mg/mL 1 μm gold particle suspension is added (in order): 5 μL DNA (1 $\mu\text{g}/\mu\text{L}$), 20 μL spermidine (0.1 M), and 50 μL CaCl_2 (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 μL 70% ethanol and resuspended in 40 μL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five μL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in

LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio- β -galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 μ L of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One μ g of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

Expression of Maize Phytyl/prenyltransferase in Soybean Somatic Embryos

Sub B9

The ability to change the levels of total tocopherol levels in plants by transforming them with sequences encoding the maize phytyl/prenyltransferase was tested by preparing transgenic soybean somatic embryos and assaying the tocopherol and oil levels. Plasmid DNA from clone poo18chste82r was used as a template for the amplification of the open reading from pcr by using the following two primers AGC GCG GCC GCA TGG ACG CGC TTC GCC TAC GGC CGT(forward primer) and AGC GCG GCC GCT CAC CGC ACC AGA GGG ATG AGC AG(reverse primer). Pfu polymerase was used according to the manufacturers recommendations (Stratagene). The following pcr reaction mix contained the following: 5ng plasmid, 25nmoles dNTPs, 5% DMSO, 1x pcr buffer (supplied), 30nmoles primers, 5U pfu polymerase in 100ul reaction volume. The pcr reaction conditions were as follows: Step 1, 45s 94°C; step 2 25 cycles of 45s 94°C, 45s 58°C annealing, 2min extension 72°C. Step 3 72°C 10min, step 4 0°C. The pcr product was purified by agarose gel electrophoresis (1% agarose in TAE), the ethidium bromide visualized band cut out and purified from the gel by using a QIAquick Gel Extraction kit (Qiagen) according to the manufacturers recommendations. The purified pcr product (200ng) was ligated into the *srf*I site of the plasmid PCR-Script cloning vector and the resultant plasmid was used to transform E.coli DH10 cells. Colonies containing the 1.2kb NotI fragment were identified by antibiotic (ampicillin selection) and blue / white (IPTG + X-gal)

selection of colonies on LB/Amp plates. White (recombinant) colonies were picked and grown overnight on liquid LB/Amp culture. Positive clones were identified by plasmid preparation and restriction digest analysis for the presence of the 1.2kB *NotI* fragment. Positive clones were used as template to fully sequence the phytyl transferase orf (both strands). Plasmids containing the correct insert verified by nucleic acid sequence were digested with *NotI* and the 1.2kb fragment ligated to *NotI*-digested and phosphatase-treated pKS67. The plasmid pKS67 was prepared by replacing in pRB20 (described in U.S. Patent No. 5,846,784) the 800 bp Nos 3' fragment, with the 285 bp Nos 3' fragment containing the polyadenylation signal sequence and described in Depicker et al. (1982) *J. Mol. Appl. Genet.* 1:561-573. Clones were screened for the sense and antisense orientation of the phytyl/prenyltransferase insert fragment by restriction enzyme digestion.

Transformation of Soybean Somatic Embryo Cultures

The stock solutions and media shown in Table 4 were used for transformation and propagation of soybean somatic embryos:

Table 4 – Stock Solution and Media

Stock Solutions	
<u>MS Sulfate 100x stock</u>	<u>(g/L)</u>
MgSO ₄ ·7H ₂ O	37.0
MnSO ₄ ·H ₂ O	1.69
ZnSO ₄ ·7H ₂ O	0.86
CuSO ₄ ·5H ₂ O	0.0025
<u>MS Halides 100x stock</u>	
CaCl ₂ ·2H ₂ O	44.0
KI	0.083
CoCl ₂ ·6H ₂ O	0.00125
KH ₂ PO ₄	17.0
H ₃ BO ₃	0.62
Na ₂ MoO ₄ ·2H ₂ O	0.025
Na ₂ EDTA	3.724
FeSO ₄ ·7H ₂ O	2.784
<u>B5 Vitamin stock</u>	
myo-inositol	100.0
nicotinic acid	1.0
pyridoxine HCl	1.0
thiamine	10.0

Media
<u>SB55 (per Liter)</u>
10 mL of each MS stock
1 mL of B5 Vitamin stock
0.8 g NH ₄ NO ₃
3.033 g KNO ₃
1 mL 2,4-D (10 mg/mL stock)
0.667 g asparagine
pH 5.7
<u>SB103 (per Liter)</u>
1 pk. Murashige & Skoog salt mixture*
60 g maltose
2 g gelrite
pH 5.7
<u>SB148 (per Liter)</u>
1 pk. Murashige & Skoog salt mixture*
60 g maltose
1 mL B5 vitamin stock
7 g agarose
pH 5.7

*(Gibco BRL)

Soybean embryonic suspension cultures were maintained in 35 mL liquid media (SB55) on a rotary shaker (150 rpm) at 28°C with a mix of fluorescent and incandescent lights providing a 16 h day 8 h night cycle. Cultures were subcultured every 2 to 3 weeks by inoculating approximately 35 mg of tissue into 35 mL of fresh liquid media.

Soybean embryonic suspension cultures were transformed with the plasmid containing the phytyl/prenyltransferase sequence (positive orientation) by the method of particle gun bombardment (see Klein et al. (1987) *Nature* 327:70-73) using a DuPont Biolistic PDS1000/He instrument. Five µL of pKS93s plasmid DNA (1 g/L), 50 µL CaCl₂ (2.5 M), and 20 µL spermidine (0.1 M) were added to 50 µL of a 60 mg/mL 1 mm gold particle suspension. The particle preparation was

agitated for 3 minutes, spun on a microfuge for 10 seconds and the supernate removed. The DNA-coated particles were then washed once with 400 μ L of 70% ethanol and resuspended in 40 μ L of anhydrous ethanol. The DNA/particle suspension was sonicated three times for 1 second each. Five μ L of the DNA-coated gold particles were then loaded on each macro carrier disk.

Approximately 300 to 400 mg of two-week-old suspension culture was placed in an empty 60 mm X 15 mm petri dish and the residual liquid removed from the tissue using a pipette. The tissue was placed about 3.5 inches away from the retaining screen and bombarded twice. Membrane rupture pressure was set at 1100 psi and the chamber was evacuated to -28 inches of Hg. Two plates were bombarded, and following bombardment, the tissue was divided in half, placed back into liquid media, and cultured as described above.

Fifteen days after bombardment, the liquid media was exchanged with fresh SB55 containing 50 mg/mL hygromycin. The selective media was refreshed weekly. Six weeks after bombardment, green, transformed tissue was isolated and inoculated into flasks to generate new transformed embryonic suspension cultures.

Transformed embryonic clusters were removed from liquid culture media and placed on a solid agar media, SB103, containing 0.5% charcoal to begin maturation. After 1 week, embryos were transferred to SB103 media minus charcoal. After 5 weeks on SB103 media, maturing embryos were separated and placed onto SB148 media. During maturation embryos were kept at 26°C with a mix of fluorescent and incandescent lights providing a 16 h day 8 h night cycle. After 3 weeks on SB148 media, embryos were analyzed for the expression of the tocopherols. Each embryonic cluster gave rise to 5 to 20 somatic embryos.

Non-transformed somatic embryos were cultured by the same method as used for the transformed somatic embryos.

Analysis of Transformed Somatic Embryos

At the end of 3 weeks on SB148 medium somatic embryos were harvested from 33 independently transformed lines. Pools of five embryos/event were pooled, the fresh weight noted, the embryos frozen on dry ice and lyophilized overnight. The corresponding dry weight was noted, the embryos pulverized with a glass rod and tocopherols and oil extracted by the addition of 0.5ml heptane (18h,

room temperature, dark). The embryos were re-extracted with 0.25ml of heptane the solutions pooled and centrifuged (5min, 12000g). The supernatant was stored in amber hplc autosampler vials at -20°C prior to analysis.

HPLC analysis of the extracts was carried out using an HP1100 system (Agilent Technologies). 25ul of the heptane sample was applied to a Lichrosphere Si 60 column (5micron 4 x 12.5mm). The column was eluted with heptane/isopropanol (98:2 v/v) at a flow rate of 1ml/min. After 6minutes all four tocopherol isomers were eluted, as detected by a HP1100 fluorescence detector (excitation wavelength 295nm, emission wavelength 330nm). Individual tocopherol standards (Matreya) were diluted with HPLC grade heptane to levels between 1 and 200ng/ul to construct a six point external standard curve. Tocopherols in each sample were quantified using a standard curve run on the same day as the samples.

Total oil content of the samples was estimated by quantitative gas chromatography of the fatty acid methyl esters. 50ul samples were derivitized by addition to 0.5ml of a 1% (v/v) solution of sodium methoxide in methanol, 1 ug of undecanoic acid (17:0) dissolved in toluene was added as an internal standard. Derivitized fatty acids were extracted in 400ul heptane, fatty acids separated by glc and the peak heights quantitated by using a HP 6890 gas liquid chromatograph equipped with a fused silica capillary column 30m x i.d. 0.25mm coated with polar phase Omegawax 320 (Supelco In, Bellfonte,, PA), autosampler, flame ionization detector and ChemStation software on a HP

The example shown in Table 5 shows the data from 33 independent transformed lines of somatic soy embryos (five pooled embryos per line) transformed with KS67 containing the maize phytyl/prenyltransferase in the positive orientation. Normal ratios of tocopherol (ngT) / oil (ugOil) in somatic embryos are 2-5. Overexpression of the phytyl/prenyltransferase has increased the amount of tocopherol relative to oil. In particular in samples 16 and 17 the ng/ugOil ratios have doubled to be 10.9 and 10.1 respectively.

Table 6 – Detailed Analysis of each of the Five Embryos in Transformed Lines 15 (control), 16, 17 and 18 (control).

Sample	Oil (mg)	Tocopherol (ng)	ngT/ugOil
SC5 15-1	1.36	10.6	7.8
SC5 15-2	0.75	7.2	9.7
SC5 15-3	0.93	5.4	5.8
SC5 15-4	6.67	37.04	5.6
SC5 15-5	1.35	10.1	7.5
SC5 16-1	0.8	15.8	19.7
SC5 16-2	0.4	10.7	26.8
SC5 16-3	4.21	27.5	6.5
SC5 16-4	0.2	3.5	17.5
SC5 16-5	2.7	35.7	13.2
SC5 17-1	0.4	44.3	111
SC5 17-2	0.2	39.6	200
SC5 17-3	5	58	11.7
SC5 17-4	1.29	11.6	9
SC5 17-5	24.7	197.8	13.5
SC5 18-1	32.1	43.4	1.4
SC5 18-2	31.6	6.1	1.9
SC5 18-3	2.99	11.6	3.9
SC5 18-4	0.7	6.1	8.7
SC5 18-5	0.5	3.4	7

The single embryo analysis in Table 6 was conducted to confirm the pooled embryo data provided in Table 5. It should also be noted that an alternative embodiment of the invention involves somatic soy embryos transformed with KS67 containing the maize phytyl/prenyltransferase in the reverse orientation.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, patent applications, and computer programs cited herein are hereby incorporated by reference.

References

- Addlesee, H. A., Gibson, L. C. D., Jensen, P. E. & Hunter, C. N. (1996). Cloning, sequencing and functional assignment of the chlorophyll biosynthesis gene, chlP, of *Synechocystis* sp. PCC 6803. FEBS Lett. 389: 126-130.
- Cai, Y. & Wolk, C. P. (1990) Use of conditionally lethal gene in *Anabena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences. J. Bacteriol. 172:3138-3145.
- Crane, F. L. & Barr, R. (1971). Determination of ubiquinones. Meth. Enzymol. 18:137-165.
- Hutson, K. G. & Threlfall, D. R. (1980). Synthesis of plastoquinone-9 and phytylplastoquinone from homogentisate in lettuce chloroplasts. Biochim. Biophys. Acta 632:630-648.
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. & Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. 3: 109-136.
- Keller, Y., Bouvier, F., d'Harlingue, A. & Camara, B. (1998). Metabolic compartmentation of plastid prenolipid biosynthesis: Evidence for the involvement of a multifunctional geranylgeranyl reductase. Eur J Biochem. 251 413-7.
- Lopez, J. C., Ryan, S. & Blankenship, R. E. (1996). Sequence of the bchG gene from *Chloroflexus aurantiacus*: relationship between chlorophyll synthase and other polyprenyltransferases. J. Bacteriol. 178: 3369-3373 .

Norris, S. R., Shen, X. & DellaPenna, D. (1998). Complementation of the Arabidopsis pds1 mutation with the gene encoding p-hydroxyphenylpyruvate dioxygenase. Plant Phys. 117:1317-1323.

Pennock, J. F. (1985). Biosynthesis of plastoquinone. Meth. Enzymol. 110:313-319.

Seffens, W., Almoguera, C., Wilde, H., Vonder Haar, R. & Thomas, T. (1990). Molecular analysis of a phylogenetically conserved carrot gene: developmental and environmental regulation. Dev. Genet. 11:65-76.

Soll, J. (1987). α -Tocopherol and plastoquinone synthesis in chloroplast membranes. Meth. Enzymol. 148:383-392.

Soll, J., Kemmerling, M. & Schultz, G. (1980). Tocopherol and plastoquinone synthesis in Spinach chloroplast subfractions. Arch. Biochem. Biophys. 204:544-550.

Williams, J. G. K. (1988). Construction of specific mutations in Photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. Meth. Enzymol. 167:766-778.

The present invention will be further described by reference to the following detailed examples. It is understood, however, that there are many extensions, variations, and modification on the basic theme of the present invention beyond that shown in the examples and description, which are within the spirit and scope of the present invention. All publications, patents, and patent application cited herein are hereby incorporated by reference.